

**Some Methods for Comparing
Heteroscedastic Regression Models**

Wu Hao

NATIONAL UNIVERSITY OF SINGAPORE

2011

Some Methods for Comparing
Heteroscedastic Regression Models

Wu Hao

(B.Sc. National University of Singapore)

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF SCIENCE
DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgements

I would like to take this opportunity to express my deepest gratitude to everyone who has provided me their support, advice and guidance throughout this thesis.

First and foremost I would like to thank my supervisor, Professor Zhang Jin Ting, for his guidance and assistance during my graduate study and research. This thesis would not have been possible without his inspiration and expertise. I would like to thank him for teaching me how to undertake researches and spending his valuable time revising this thesis.

I would also like to express my sincere gratitude to my family and friends for their help in completing this thesis.

List of Tables

4.1	Parameter configurations for simulations	30
4.2	Empirical sizes and powers for 2-sample test($p = 5$).	32
4.3	Empirical sizes and powers for 2-sample test($p = 10$).	33
4.4	Empirical sizes and powers for 2-sample test($p = 20$).	34
4.5	Empirical sizes and powers for 3-sample test($p = 10$).	35
4.6	Empirical sizes and powers for 5-sample test($p = 2$).	37
4.7	Empirical sizes and powers for 5-sample test($p = 5$).	38
4.8	Test Results	39
4.9	Test Results	41

Contents

Acknowledgements	iii
List of Tables	iv
Abstract	vii
1 Introduction	1
1.1 Motivation	1
1.2 Organization of this Thesis	3
2 Literature Review	4
2.1 Chow's Test For Homogeneous Variances	4
2.2 Toyoda's Modified Chow Test	7
2.3 Jayatissa's exact small sample test and Watt's Wald Test	9
2.4 Conerly and Mansfield's Approximate Test	11
3 Models and Methodology	15
3.1 Generalization of Chow's Test	15

3.2	Wald-type Test	19
4	Simulation and Real Life Example	28
4.1	Simulation	28
4.2	Real Life Examples	39
5	Conclusion and Future Research	43
6	Appendix	45
6.1	Matlab code in common for simulations	45
6.2	Simulation Studies for 2-sample and k-sample cases	51
	Bibliography	58

Abstract

The Chow's test was proposed to test the equality of coefficients of two linear regression models under the assumption of homogeneity. We generalize the Chow's test and modified Chow's test to k -sample case. We propose the Wald-type test for testing the equivalence of coefficients of k linear models under homogeneity assumption. For heteroscedastic case, we adopt the same Wald test statistic using the approximate degree of freedom method to approximate its distribution. The simulation studies and real life examples are presented to examine the performances of the proposed test statistics.

Keywords: linear models; Chow's test; heteroscedasticity; approximate degrees of freedom test; Wald statistic

Chapter 1

Introduction

In econometrics, the linear regression model has often been widely applied to the measurement of econometric relationships. The data used for analysis have been collected over a period of time, therefore the question often arises as to whether the same relationship remains stable in two periods of time, for example, pre-World War II and post-World World II. Statistically, this question can be simplified to test whether subsets of coefficients in two regressions are equal.

1.1 Motivation

A pioneering work in this research field was done by Chow (1960), in which he proposed a F-test to conduct the hypothesis testing. Chow's test is proposed under the homogeneous variance assumption and it is proved that the test works well as long as at least one of the sample sizes is large. However, when the error

terms of the two models differ, the Chow's test may provide inaccurate results. Toyoda (1974) and Schmidt and Sickles (1977) have demonstrated that the presence of heteroscedasticity can lead to serious distortions in the size of the test. Two alternative tests for equality of coefficients under heteroscedasticity have been proposed by Jayatissa (1977) and Watt (1979). Jayatissa proposed an exact small sample test and Watt developed an asymptotic Wald test. Both of these tests have their drawbacks and hence Ohtani and Toyoda (1985) investigated the effects of increasing the number of regressors on the small sample properties of these two tests and found that the Jayatissa test cannot always be applied. Gurland and McCullough (1962) proposed the two-stage test which consists of pre-test for equality of variances and the main-test for equality of means. Ohtani and Toyoda (1986) extended the analysis to the case of a general linear regression. Other alternative testing procedures include Ali and Silver's (1985) two approximate test based on Pearson system using the moments of the statistics under the null hypothesis, Moreno, Torres and Casella's (2005) Bayesian approaches, and Conerly and Mansfield's (1988, 1989) approximation test which can be implemented easily.

To this end, we may notice that in reality most of the time we are dealing with the problem under heteroscedasticity and it is also very likely to encounter problems that involve more than two samples. In this paper, we will propose methods which are intended to comprise the disadvantages, at least a few, of the

methods mentioned above under the condition of heteroscedasticity. The desirable method would be easily implemented and generalized to k sample cases.

1.2 Organization of this Thesis

The remaining part of this thesis is organized as follows: firstly we will review the existing methods to test the equivalence of coefficients of two linear models in Chapter 2. In the first two sections of Chapter 3, we generalize the existing methods to problems of k -sample case and it is followed by our proposal of an approximate degrees of freedom (ADF) test of the Wald statistics in the following section of Chapter 3. In Chapter 4 we present simulation results and real life examples. Finally, we summarize and conclude the thesis in Chapter 5.

Chapter 2

Literature Review

Testing for equality of regression coefficients in different populations is widely used in econometric and other research. It dates back to 1960's when Chow proposed a testing statistic to compare the coefficients of two linear models under the assumption of homogeneity. In practice, however, the homogeneous assumption can rarely hold, therefore various modified statistics based on Chow's test have been formulated. A brief literature review is given in the following sections.

2.1 Chow's Test For Homogeneous Variances

2.1.1 Chow's Test For Two Sample Cases

Assume that we have two independent linear regression models based on n_1 and n_2 observations, namely,

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, i = 1, 2, \quad (2.1)$$

where \mathbf{Y}_i is an $n_i \times 1$ vector of observations, \mathbf{X}_i is an $n_i \times p$ matrix of observed values on the p explanatory variables, and $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ vector of error terms. The errors are assumed to be independent normal random variables with zero mean and variances σ_1^2 and σ_2^2 , respectively. Then to test the equivalence of two sets of coefficient vectors, the hypothesis can be stated formally as

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 \text{ versus } H_A : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2. \quad (2.2)$$

Under H_0 , the model can be combined as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.3)$$

where $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}$ and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{n_2} \end{bmatrix} \quad (2.4)$$

Denote the error sum of squares for this model by

$$\mathbf{e}^T \mathbf{e} = \mathbf{Y}^T [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{Y} \quad (2.5)$$

It can be further written as

$$\mathbf{e}^T \mathbf{e} = \boldsymbol{\epsilon}^T [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T [\mathbf{I} - \mathbf{P}_X] \boldsymbol{\epsilon} \quad (2.6)$$

where $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ denotes the “hat” matrix of \mathbf{X} in (2.3).

Under H_A , the model may be written as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\epsilon} = \mathbf{X}^* \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \boldsymbol{\epsilon} \quad (2.7)$$

The sum of squared errors for each model is

$$\mathbf{e}_i^T \mathbf{e}_i = \mathbf{Y}_i^T [\mathbf{I} - \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T] \mathbf{Y}_i = \mathbf{Y}_i^T [\mathbf{I} - \mathbf{P}_{X_i}] \mathbf{Y}_i, i = 1, 2, \quad (2.8)$$

where $\mathbf{P}_{X_i} = \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T$ denotes the “hat” matrix for data set $i = 1, 2$. The sum of the squared errors for the unrestricted case becomes

$$\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2 = \mathbf{Y}_1^T (\mathbf{I} - \mathbf{P}_{X_1}) \mathbf{Y}_1 + \mathbf{Y}_2^T (\mathbf{I} - \mathbf{P}_{X_2}) \mathbf{Y}_2 = \mathbf{Y}^T [\mathbf{I} - \mathbf{P}_{X^*}] \mathbf{Y} \quad (2.9)$$

where

$$\mathbf{P}_{X^*} = \begin{bmatrix} \mathbf{P}_{X_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{X_2} \end{bmatrix}, \quad (2.10)$$

since $(\mathbf{I} - \mathbf{P}_{X^*}) \mathbf{X}^* = \mathbf{0}$, $\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2 = \boldsymbol{\epsilon}^T [\mathbf{I} - \mathbf{P}_{X^*}] \boldsymbol{\epsilon}$.

The test statistic for the Chow test is

$$F = \frac{[\mathbf{e}^T \mathbf{e} - \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2]/p}{[\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2]/(n_1 + n_2 - 2p)}. \quad (2.11)$$

Using the notation introduced above this can be written as

$$F = \frac{\boldsymbol{\epsilon}^T (\mathbf{P}_{X^*} - \mathbf{P}_X) \boldsymbol{\epsilon} / p}{\boldsymbol{\epsilon}^T [\mathbf{I} - \mathbf{P}_{X^*}] \boldsymbol{\epsilon} / (n_1 + n_2 - 2p)}, \quad (2.12)$$

which is a ratio of quadratic forms. The independence of the numerator and denominator in F follows since

$$(\mathbf{P}_{X^*} - \mathbf{P}_X) \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{P}_{X^*}) = \mathbf{0}. \quad (2.13)$$

If $\sigma_1^2 = \sigma_2^2$, then the statistic in (2.12) has an F distribution with p and $n_1 + n_2 - 2p$ degrees of freedom under the null hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. However, in practice

the condition of homogeneous variance does not always hold and it is more common to deal with heteroscedastic variance data. Hence the problem becomes how to test two sets of coefficients under heteroscedasticity.

2.2 Toyoda's Modified Chow Test

When the homogeneity assumption is invalid, the usual F -test may result in misleading conclusions. Since F is the ratio of independent quadratic forms, we may apply Satterthwaite's approximation to the numerator and denominator as suggest by Toyoda (1974). For the term $(\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2)$ of the denominator in (2.10), Toyoda approximated it as a scalar multiple of a chi-square distribution. The scalar multiple and the number of degrees of freedom are chosen so as to make the first two moments of the exact distribution and the approximate distribution equal. As for the numerator, Toyoda noted that

$$\mathbf{e}^T \mathbf{e} - \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2 = \boldsymbol{\epsilon}^T (\mathbf{P}_{X^*} - \mathbf{P}_X) \boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T \mathbf{A} \boldsymbol{\epsilon} \quad (2.14)$$

where \mathbf{A} is an idempotent matrix with rank p . If $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$, then $\boldsymbol{\epsilon}^T \mathbf{A} \boldsymbol{\epsilon}$ would be distributed as $\sigma^2 \chi_p^2$. However since

$$\text{cov}(\boldsymbol{\epsilon}) = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_{n_1} \end{bmatrix} \quad (2.15)$$

Toyoda presumed that the distribution of $(\mathbf{e}^T \mathbf{e} - \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2)$ can be approximated by $\sigma^2 \chi_p^2$, where “ σ^2 is any well-chosen weighted average of σ_1^2 and σ_2^2 ” and p is fixed

to be the same as the degree of freedom of Chow's test.

Toyoda showed that the denominator of (2.10) may be approximated by $a_2\chi_{(f_2)}^2$, where

$$a_2 = \frac{(n_1 - p)\sigma_1^4 + (n_2 - p)\sigma_2^4}{(n_1 - p)\sigma_1^2 + (n_2 - p)\sigma_2^2} \quad (2.16)$$

and

$$f_2 = \frac{[(n_1 - p)\sigma_1^2 + (n_2 - p)\sigma_2^2]^2}{(n_1 - p)\sigma_1^4 + (n_2 - p)\sigma_2^4} \quad (2.17)$$

For convenience, Toyoda chose the weighted average of σ_1^2 and σ_2^2 to be $(n_1 - p)\sigma_1^2/[(n_1 - p)\sigma_1^2 + (n_2 - p)\sigma_2^2]$ and $(n_2 - p)\sigma_2^2/[(n_1 - p)\sigma_1^2 + (n_2 - p)\sigma_2^2]$ respectively, and in consequence $\sigma^2 = a_2$. Then for the test statistic

$$F^* = \frac{[\mathbf{e}^T \mathbf{e} - (\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2)]/p}{(\mathbf{e}_1^T \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_2)/f_2}, \quad (2.18)$$

the numerator and the denominator are distributed as $pa_2\chi_{(p)}^2$ and $f_2a_2\chi_{(f_2)}^2$, respectively. Hence F^* is approximately distributed as $F(p, f_2)$ and the approximate distribution of F is $((n_1 + n_2 - 2p)/f_2)F(p, f_2)$ rather than $F(p, n_1 + n_2 - 2p)$.

Schmidt and Sickles (1977) examined Toyoda's approximation and found that the true significant level of the test was actually quite different from the Toyoda's approximation in many cases. They concluded that the approximation of the denominator is reasonable and is not apt to be the major source of inaccuracy. However, Toyoda's approximation for numerator depends only on n_1, n_2, p , and $\theta = \sigma_2^2/\sigma_1^2$, whereas the exact distribution depends on the form of X_1 and X_2 as well as n_1, n_2, k and $\theta = \sigma_2^2/\sigma_1^2$. According to the simulation result, they questioned

Toyoda's assertion that "if at least one of the two samples is of large size, the Chow test is robust for any finite variations of variance" and concluded that increasing one sample size does not necessarily increase the reliability of Toyoda's test. Based on their numerical result and theoretical proof, they also concluded that Toyoda's approximation was better when the error variances are approximately equal, but that it may not be very good if the variances differ greatly.

2.3 Jayatissa's exact small sample test and Watt's

Wald Test

In 1970's, two alternative tests for equality of coefficients under heteroscedasticity have subsequently been proposed by Jayatissa (1977) and Watt (1979). Jayatissa suggested a small sample exact test and this test is defined as follows:

Let $\mathbf{M}_i = \mathbf{I}_{n_i} - \mathbf{H}_i = \mathbf{Z}_i \mathbf{Z}_i^T$, where $\mathbf{Z}_i^T \mathbf{X}_i = \mathbf{0}$ and $\mathbf{Z}_i^T \mathbf{Z}_i = \mathbf{I}_{n_i-p}$. Then

$$\mathbf{d} = \hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2 \sim N(\mathbf{v}, \boldsymbol{\Sigma}), \quad (2.19)$$

where $\mathbf{v} = \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma} = \sigma_1^2(\mathbf{X}_1^T \mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2^T \mathbf{X}_2)^{-1}$.

Now consider $\mathbf{e}_i^* = \mathbf{Z}_i^T \mathbf{e}_i$ and note that

$$\mathbf{e}_i^* \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}_{n_i-p}), \quad (2.20)$$

$i = 1, 2$. Let r be the largest integer no larger than $\min((n_1 - p)/p, (n_2 - p)/p)$

and partition each \mathbf{e}_i^* into r subvectors $\mathbf{e}_{i(1)}^*, \mathbf{e}_{i(2)}^*, \dots, \mathbf{e}_{i(r)}^*$, each subvector having

p elements. Now let \mathbf{Q}_i be a $p \times p$ matrix such that $\mathbf{Q}_i^T \mathbf{Q}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1}$. Then

$$\boldsymbol{\eta}_j = \mathbf{Q}_1^T \mathbf{e}_{1(j)}^* + \mathbf{Q}_2^T \mathbf{e}_{2(j)}^*, j = 1, 2, \dots, r, \quad (2.21)$$

are mutually independent vectors, each distributed as $N(\mathbf{0}, \boldsymbol{\Sigma})$ and independently of \mathbf{d} . To this end, we have

$$\frac{\mathbf{d}^T \mathbf{S}^{-1} \mathbf{d}}{r} \frac{r - p + 1}{p}, \quad (2.22)$$

which is distributed as non-central F with p and $r - p + 1$ degrees of freedom and non-centrality parameter $\mathbf{v}^T \boldsymbol{\Sigma}^{-1} \mathbf{v}$, provided that $r \gg k$, where $\mathbf{S} = \frac{1}{r} \sum_{j=1}^r \boldsymbol{\eta}_j \boldsymbol{\eta}_j^T$.

This test is proven to be inefficient because confidence intervals based on the test statistic are large (Weerahandi, 1987). It also have few degrees of freedom and it have been to shown to lack some desirable invariance properties (Tsurumi, 1984). Jayatissa pointed out in his paper that “the degress of freedom may be very small and further research is needed to decide whether this exact small sample test is operationally superior to an approximate test based on asymptotic theory”. One possible solution would be the Wald test which was suggested by Watt (1977). The Wald test statistic can be written as:

$$C = (\hat{\beta}_1 - \hat{\beta}_2)^T (\hat{\sigma}_1^2 (\mathbf{X}_1^T \mathbf{X}_1)^{-1} + \hat{\sigma}_2^2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1})^{-1} (\hat{\beta}_1 - \hat{\beta}_2), \quad (2.23)$$

and its asymptotic distribution is χ_p^2 . Simulation studies in Watt and in Honda (1982) revealed that the Wald test is preferable to the Jayatissa test when two sample size are moderate or large. But no firm conclusions can be drawn when sample sizes are small. The problem of the Wald test is that the size of the test

is not exactly guaranteed in small samples, whereas the difficulty of the Jayatissa test lies in its low power of the test.

2.4 Conerly and Mansfield's Approximate Test

Ali and Silver (1985) proposed two approximate tests for comparing heteroscedastic regression models. One test is based on the usual F test and the other is based on a likelihood ratio test for the unequal variance case. Their results confirmed Schmidt and Sickles' assessment of Toyoda's approximation and they concluded that the standard F statistic is more robust and the difference in power for the two tests is inconsequential for many design configurations. Based on Ali and Silver's result, Conerly and Mansfield (1987) proposed an approximate test based on the same statistic as Toyoda's by using Satterthwaite's (1946) approximation not just for the numerator but also for the denominator of the usual F statistic.

The approximation of the denominator is the same as Toyoda's test, which is $a_2\chi_{(f_2)}^2$. The numerator can be treated similarly by equating its first two moments with those of $a_1\chi_{(f_1)}^2$. The resulting constants are

$$a_1 = \sum [(1 - \lambda_i)\sigma_1^2 + \lambda_i\sigma_2^2]^2 / \sum [(1 - \lambda_i)\sigma_1^2 + \lambda_i\sigma_2^2] \quad (2.24)$$

and

$$f_1 = \sum [(1 - \lambda_i)\sigma_1^2 + \lambda_i\sigma_2^2]^2 / \sum [(1 - \lambda_i)\sigma_1^2 + \lambda_i\sigma_2^2], \quad (2.25)$$

where λ_i denotes the i -th eigenvalue of $\mathbf{W} = \mathbf{X}_1^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1 + \mathbf{X}_2^T \mathbf{X}_2)^{-1}$. Combining this with the previous results, the approximate distribution of the F statistic is

$$F \sim \left(\frac{n_1 + n_2 - 2p}{p} \right) \cdot \frac{a_1 f_1}{a_2 f_2} \cdot F_{f_1, f_2}. \quad (2.26)$$

The key difference between this and Toyoda's approximation is that Toyoda used p degrees of freedom for the numerator and chose the multiplier, a_1 , to be equal to a_2 . It should be noticed that a_1 and f_1 depend on the matrices \mathbf{X}_1 and \mathbf{X}_2 through the eigenvalues λ_i . One of the shortcomings of Toyoda's approximation as pointed out by Schmidt and Sickles was that it did not incorporate the form of the \mathbf{X}_i matrices. They also demonstrated that as n_1 or $n_2 \rightarrow \infty$ with the other sample size fixed, the Toyoda's approximation differs from the actual distribution of the statistic F . The procedure proposed by Conerly and Mansfield is relatively easy to apply in practice, although the eigenvalues of $\mathbf{W} = \mathbf{X}_1^T \mathbf{X}_1 (\mathbf{X}_1^T \mathbf{X}_1 + \mathbf{X}_2^T \mathbf{X}_2)^{-1}$ and estimates of the unknown parameters σ_1^2 and σ_2^2 must be determined.

Taking this as the starting point, Conerly and Mansfield (1988) further developed a test which introduced an alternative denominator to afford a better approximation. For the modified Chow statistic, C^* is constructed by using $\theta_1 \hat{\sigma}_1^2 + \theta_2 \hat{\sigma}_2^2$ as the denominator, where θ_1 and θ_2 are constants chosen to improve the approximation. They equated the moments of the denominator,

$$E[\theta_1 \hat{\sigma}_1^2 + \theta_2 \hat{\sigma}_2^2] = \theta_1 \sigma_1^2 + \theta_2 \sigma_2^2 \quad (2.27)$$

and

$$\text{Var}(\theta_1 \hat{\sigma}_1^2 + \theta_2 \hat{\sigma}_2^2) = 2[\theta_1^2 \sigma_1^4 / (n_1 - p) + \theta_2^2 \sigma_2^4 / (n_2 - p)] \quad (2.28)$$

to the moments of $a_2 \chi_{(f_2)}^2$, yielding

$$a_2 = \frac{\theta_1^2 \sigma_1^2 / (n_1 - p) + \theta_2^2 \sigma_2^2 / (n_2 - p)}{\theta_1 \sigma_1^2 + \theta_2 \sigma_2^2} \quad (2.29)$$

and

$$f_2 = \frac{(\theta_1 \sigma_1^2 + \theta_2 \sigma_2^2)^2}{(\theta_1 \sigma_1^2)^2 / (n_1 - p) + (\theta_2 \sigma_2^2)^2 / (n_2 - p)}. \quad (2.30)$$

The first two moments of the numerator can be equated to those of $a_1 \chi_{(f_1)}^2$. The resulting constants a_1 and f_1 remain the same as (2.24) and (2.25). The test statistic C^* can be expressed as

$$C^* = \frac{(\mathbf{e}^T \mathbf{e} - \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2) / p}{\theta_1 \hat{\sigma}_1^2 + \theta_2 \hat{\sigma}_2^2} \quad (2.31)$$

. Combining this with the previous results, the approximated distribution of C^* is

$$C^* \sim (a_1 f_1 / a_2 f_2 p) F_{(f_1, f_2)}, \quad (2.32)$$

We may notice the degrees of freedom f_1 and f_2 of C^* change slowly with respect to changes in variance ratio σ_1^2 / σ_2^2 , therefore, f_1 and f_2 will not have substantial effect on the test significance level as the variance ratio changes. We have to minimize the rate of the change of the multiplier $a_1 f_1 / a_2 f_2 p$ in order to stabilize the approximation. To this end, Conerly and Mansfield suggested that $\theta_1 = (1 - \bar{\lambda})$ and $\theta_2 = \bar{\lambda}$, since

$$\begin{aligned} a_1 f_1 / a_2 f_2 p &= \sum [(1 - \lambda_i) \sigma_1^2 + \lambda_i \sigma_2^2] / (\theta_1 \sigma_1^2 + \theta_2 \sigma_2^2) \\ &= [(1 - \bar{\lambda}) \sigma_1^2 + \bar{\lambda} \sigma_2^2] / (\theta_1 \sigma_1^2 + \theta_2 \sigma_2^2), \end{aligned} \quad (2.33)$$

and this will be unity when taking the suggested value of θ_1 and θ_2 .

The resulting test statistic,

$$C^* = \frac{(\mathbf{e}^T \mathbf{e} - \mathbf{e}_1^T \mathbf{e}_1 - \mathbf{e}_2^T \mathbf{e}_2)/p}{(1 - \bar{\lambda})\hat{\sigma}_1^2 + \bar{\lambda}\hat{\sigma}_2^2}, \quad (2.34)$$

and it follows an approximate F -distribution with degrees of freedom

$$f_1^* = \{p[(1 - \bar{\lambda})\hat{\sigma}_1^2 + \bar{\lambda}\hat{\sigma}_2^2]\}^2 / \sum \{(1 - \lambda_i)\hat{\sigma}_1^2 + \lambda_i\hat{\sigma}_2^2\}^2 \quad (2.35)$$

and

$$f_2^* = \{(1 - \bar{\lambda})\hat{\sigma}_1^2 + \bar{\lambda}\hat{\sigma}_2^2\}^2 / \{[(1 - \bar{\lambda})\hat{\sigma}_1^2]^2/(n_1 - p) + [\bar{\lambda}\hat{\sigma}_2^2]^2/(n_2 - p)\}. \quad (2.36)$$

This method is easy to implement and later we will discuss the impact of this estimation on the approximation as we compare several testing methods.

Chapter 3

Models and Methodology

In this chapter, firstly we would like to generalize the methods mentioned previously to k -sample cases, where $k > 2$. Then we will propose a Wald-type statistic for testing 2-sample and k -sample cases.

3.1 Generalization of Chow's Test

3.1.1 Generalized Chow's Test For

Although it has not been mentioned in Chow's paper to generalize his method to more than two sample cases, this can be simply done through the following procedures.

Under the null hypothesis, the model can be written as:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_k \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_k \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.1)$$

where $\boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$, $N = n_1 + \cdots + n_k$. The sum of squared errors for this reduced model is

$$SSE_R = \mathbf{Y}^T [\mathbf{I} - \mathbf{P}_X] \mathbf{Y}, \quad (3.2)$$

where \mathbf{P}_X is the projection matrix of \mathbf{X} , i.e., $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Under the alternative hypothesis, the full model can be expressed as

$$\mathbf{Y} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_k \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \boldsymbol{\epsilon} = \mathbf{X}^* \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \boldsymbol{\epsilon}. \quad (3.3)$$

The sum of squared errors of this model is $SSE_F = \mathbf{Y}^T [\mathbf{I} - \mathbf{P}_{\mathbf{X}^*}] \mathbf{Y}$, where $\mathbf{P}_{\mathbf{X}^*}$ is the projection matrix of \mathbf{X}^* . It follows that the Chow-type test for more than two sample cases is

$$\mathbf{T} = \frac{(SSE_R - SSE_F)/[(k-1)p]}{SSE_F/(N - kp)}, \quad (3.4)$$

where $(k-1)p$ is the difference of the degree of freedom between these two models.

The test statistic can be simplified to

$$\mathbf{T} = \frac{\mathbf{Y}^T (\mathbf{P}_{\mathbf{X}^*} - \mathbf{P}_X) \mathbf{Y} / (k-1)p}{\mathbf{Y}^T (\mathbf{I} - \mathbf{P}_{\mathbf{X}^*}) \mathbf{Y} / (N - kp)} \quad (3.5)$$

and under the homogeneous variance condition, this test statistic follows F distribution with $(k - 1)p$ and $N - kp$ degrees of freedom.

3.1.2 Generalized Modified Chow's Test

We may notice that the fundamental idea of the modified Chow test, for example Toyoda's test and Conerly and Mansfield's test, is to use χ^2 approximation matching the first two moments of the F -type test statistics. Since these two methods have not been generalized to the k -sample case, in this section we will construct a modified Chow test statistic for k -sample cases based on the same methodology.

The numerator of generalized Chow's test in (3.5) is $\mathbf{Y}^T(\mathbf{P}_{\mathbf{X}^*} - \mathbf{P}_{\mathbf{X}})\mathbf{Y}$, as the degree of freedom part is omitted for simplicity. Let \mathbf{Q} denote $\mathbf{P}_{\mathbf{X}^*} - \mathbf{P}_{\mathbf{X}}$, then we have $\mathbf{Y}^T\mathbf{Q}\mathbf{Y}$, where \mathbf{Q} is an idempotent matrix. Then $\mathbf{Y}^T\mathbf{Q}\mathbf{Y}$ can be further expressed as

$$\mathbf{Y}^T\mathbf{Q}\mathbf{Y} = \mathbf{Y}^T\mathbf{Q}^2\mathbf{Y} = \mathbf{Z}^T\mathbf{D}^{1/2}\mathbf{Q}\mathbf{Q}^T\mathbf{D}^{1/2}\mathbf{Z} = \mathbf{Z}^T\mathbf{A}\mathbf{Z} \quad (3.6)$$

where \mathbf{Z} follows the standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{I}_N)$; \mathbf{D} is a diagonal matrix with diagonal entries to be $(\sigma_1^2\mathbf{I}_{n_1}, \dots, \sigma_k^2\mathbf{I}_{n_k})$; and $\mathbf{A} = \mathbf{D}^{1/2}\mathbf{Q}\mathbf{Q}^T\mathbf{D}^{1/2}$. Now if we

decompose \mathbf{Q} into k blocks of size $n_i \times N$ each so that $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_k]$, then

$$\mathbf{A} = \mathbf{D}^{1/2}(\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_k) \begin{pmatrix} \mathbf{Q}_1^T \\ \mathbf{Q}_2^T \\ \vdots \\ \mathbf{Q}_k^T \end{pmatrix} \mathbf{D}^{1/2} = \sigma_1^2(\mathbf{Q}_1 \mathbf{Q}_1^T) + \dots + \sigma_k^2(\mathbf{Q}_k \mathbf{Q}_k^T) \quad (3.7)$$

It can be proved that the quadratic term $\mathbf{Z}^T \mathbf{A} \mathbf{Z}$ can be approximated by a χ^2 distribution $a_1 \chi_{f_1}^2$ by matching the first two moments, where the scalar multiplier a_1 equals to $\frac{\text{tr}(\mathbf{A}^2)}{\text{tr}(\mathbf{A})}$ and the degrees of freedom f_1 equals to $\frac{\text{tr}^2(\mathbf{A})}{\text{tr}(\mathbf{A}^2)}$.

Using similar ideas as Conerly and Mansfield's, if we equate the first moments of the numerator and the denominator, the multiple scalars of F distribution can be cancelled out. Let $\mathbf{S} = \sum_{i=1}^k \hat{\sigma}_i^2 \text{tr}(\mathbf{Q}_i^T \mathbf{Q}_i)$ and it should be noticed that

$$\mathbb{E}(\mathbf{Z}^T \mathbf{A} \mathbf{Z}) = \mathbb{E}(\mathbf{S}) = \sum_{i=1}^k \sigma_i^2 \text{tr}(\mathbf{Q}_i^T \mathbf{Q}_i). \quad (3.8)$$

Since the equivalence of their expectations holds, taking \mathbf{S} as the denominator of the test statistic will greatly simplify the computation. \mathbf{S} takes the form of $\theta_1 \hat{\sigma}_1^2 + \theta_2 \hat{\sigma}_2^2 + \dots + \theta_k \hat{\sigma}_k^2$, therefore it can be approximated by a χ^2 distribution.

We can generalize the formula (2.29) to k -sample case to calculate its degrees of freedom.

The modified Chow's test for multi-sample case can be constructed as

$$\mathbf{T} = \frac{\mathbf{Y}^T \mathbf{Q} \mathbf{Y}}{\sum_{i=1}^k \hat{\sigma}_i^2 \text{tr}(\mathbf{Q}_i^T \mathbf{Q}_i)} \quad (3.9)$$

then \mathbf{T} follows F_{f_1, f_2} distribution, where

$$f_1 = \frac{\text{tr}^2(\sum_{i=1}^k \sigma_i^2 \mathbf{Q}_i^T \mathbf{Q}_i)}{\text{tr}((\sum_{i=1}^k \sigma_i^2 \mathbf{Q}_i^T \mathbf{Q}_i)^2)} \quad (3.10)$$

and

$$f_2 = \frac{\text{tr}^2(\sum_{i=1}^k \sigma_i^2 \mathbf{Q}_i^T \mathbf{Q}_i)}{\sum_{i=1}^k \sigma_i^4 \text{tr}^2(\mathbf{Q}_i^T \mathbf{Q}_i) / (n_i - p)} \quad (3.11)$$

We will examine and compare the performance of test statistics mentioned previously via simulation studies in Chapter 4.

3.2 Wald-type Test

Assume that we have two independent linear regression models based on n_1 and n_2 observations, namely, $\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, i = 1, 2$, where \mathbf{Y}_i is an $n_i \times 1$ vector of observations, \mathbf{X}_i is an $n_i \times k$ matrix of observed values on the k explanatory variables, and $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ vector of error terms. The errors are assumed to have $\boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma_i^2 \mathbf{I}_{n_i}), i = 1, 2$. We want to test the equivalence of two sets of coefficient vectors: $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ versus $H_A : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$. The problem can also be expressed as

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}, \text{ vs } H_A : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}, \quad (3.12)$$

where $\mathbf{C} = \begin{bmatrix} \mathbf{I}_p & -\mathbf{I}_p \end{bmatrix}_{p \times 2p}$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T]^T$. Note that this is a special case of general linear hypothesis testing (GLHT) problem : $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{c}$, vs $H_A : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{c}$ with $\mathbf{c} = \mathbf{0}$. The GLHT problem is very general as it includes all the contrasts that we are interested to test. For example, when the test $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is rejected, it is of interest to test further, e.g., if $\boldsymbol{\beta}_1 = 2\boldsymbol{\beta}_2$. The testing problem can be written in the form of (3.12) with $\mathbf{C} = [\mathbf{I}_p, -2\mathbf{I}_p]_{p \times 2p}$. Therefore the Wald-type test can be implemented in more general testing problems.

For $i = 1, 2$, the ordinary least squares estimator of $\boldsymbol{\beta}_i$ and the unbiased estimator of σ_i^2 are

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{Y}_i \text{ and } \hat{\sigma}_i^2 = (n_i - p)^{-1} \mathbf{Y}_i^T (\mathbf{I}_{n_i} - \mathbf{X}_i (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T) \mathbf{Y}_i. \quad (3.13)$$

Moreover, we have

$$\hat{\boldsymbol{\beta}}_i \sim N_p(\boldsymbol{\beta}_i, \sigma_i^2 (\mathbf{X}_i^T \mathbf{X}_i)^{-1}), \hat{\sigma}_i^2 \sim \sigma_i^2 \chi_{n_i-p}^2 / (n_i - p). \quad (3.14)$$

Let $\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_1^T, \hat{\boldsymbol{\beta}}_2^T]^T$, then it is an unbiased estimator of $\boldsymbol{\beta}$ and we have $\hat{\boldsymbol{\beta}} \sim N_{2p}(\boldsymbol{\beta}, \boldsymbol{\Sigma}_\beta)$, where $\boldsymbol{\Sigma}_\beta = \text{diag}[\sigma_1^2 (\mathbf{X}_1^T \mathbf{X}_1)^{-1}, \sigma_2^2 (\mathbf{X}_2^T \mathbf{X}_2)^{-1}]$. It follows that

$$\mathbf{C}\hat{\boldsymbol{\beta}} \sim N_p(\mathbf{C}\boldsymbol{\beta}, \mathbf{C}\boldsymbol{\Sigma}_\beta \mathbf{C}^T). \quad (3.15)$$

This suggests that for testing the problem (3.12), we can use the following Wald-type test statistic

$$T = (\mathbf{C}\hat{\boldsymbol{\beta}})^T (\mathbf{C}\hat{\boldsymbol{\Sigma}}_\beta \mathbf{C}^T)^{-1} \mathbf{C}\hat{\boldsymbol{\beta}}, \quad (3.16)$$

where $\hat{\Sigma}_\beta = \text{diag}[\hat{\sigma}_1^2(\mathbf{X}_1^T \mathbf{X}_1)^{-1}, \hat{\sigma}_2^2(\mathbf{X}_2^T \mathbf{X}_2)^{-1}]$.

It should be noticed that due to the generalized form of Wald's statistics, it can be easily extended to k -sample cases. Let

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_p & \cdots & \mathbf{0}_p & -\mathbf{I}_p \\ \mathbf{0}_p & \mathbf{I}_p & \cdots & \mathbf{0}_p & -\mathbf{I}_p \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_p & \mathbf{0}_p & \cdots & \mathbf{I}_p & -\mathbf{I}_p \end{bmatrix}_{q \times kp}, \quad (3.17)$$

where $q = (k - 1)p$, then the hypothesis testing for equivalence of the coefficients of k linear regression models can be expressed as

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}, \text{ vs } H_A : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{0}. \quad (3.18)$$

and the Wald-type statistic is the same as equation (3.16), where

$$\hat{\Sigma}_\beta = \text{diag}[\hat{\sigma}_1^2(\mathbf{X}_1^T \mathbf{X}_1)^{-1}, \dots, \hat{\sigma}_k^2(\mathbf{X}_k^T \mathbf{X}_k)^{-1}].$$

3.2.1 F-test of Homogeneous Regression Models

When the homogeneity assumption of σ_1^2 and σ_2^2 holds, i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$, it is natural to estimate σ^2 by their pooled estimator $\hat{\sigma}_{pool}^2 = \sum_{i=1}^2 (n_i - p) \hat{\sigma}_i^2 / (n_1 + n_2 - 2p)$. Let $\mathbf{D} = \text{diag}[(\mathbf{X}_1^T \mathbf{X}_1)^{-1}, (\mathbf{X}_2^T \mathbf{X}_2)^{-1}]$, then under the variance homogeneity assumption so that Σ_β can be estimated by $\hat{\sigma}_{pool}^2 \mathbf{D}$, it is easy to see that

$$T/p = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T (\mathbf{C}\mathbf{D}\mathbf{C}^T)^{-1} \mathbf{C}\hat{\boldsymbol{\beta}} / (p\sigma^2)}{\hat{\sigma}_{pool}^2 / \sigma^2} \sim F_{p, n_1 + n_2 - 2p}, \quad (3.19)$$

Therefore, when the variance homogeneity assumption is valid, a usual F -test can be used to test the GLHT problem(3.14). This test statistic can be simply generalized to k -sample cases,i.e.,

$$T/q = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T(\mathbf{C}\mathbf{D}\mathbf{C}^T)^{-1}\mathbf{C}\hat{\boldsymbol{\beta}}/(q\sigma^2)}{\hat{\sigma}_{pool}^2/\sigma^2} \sim F_{q,(N-kp)}, \quad (3.20)$$

where $N = n_1 + n_2 + \cdots + n_k$. In practice, however, the homogeneity assumption is often violated so that the above F -test is no longer valid. A new test should be proposed.

3.2.2 ADF-test for Heteroscedastic Regression Models

Here we shall propose an ADF test which is obtained via properly modifying the degrees of freedom of Wald's statistics. For this end, we set

$$\mathbf{z} = (\mathbf{C}\boldsymbol{\Sigma}_\beta\mathbf{C}^T)^{-\frac{1}{2}}\mathbf{C}\hat{\boldsymbol{\beta}}, \mathbf{W} = (\mathbf{C}\boldsymbol{\Sigma}_\beta\mathbf{C}^T)^{-\frac{1}{2}}\mathbf{C}\hat{\boldsymbol{\Sigma}}_\beta\mathbf{C}^T(\mathbf{C}\boldsymbol{\Sigma}_\beta\mathbf{C}^T)^{-\frac{1}{2}} \quad (3.21)$$

so that equivalently we can write

$$T = \mathbf{z}^T\mathbf{W}^{-1}\mathbf{z}. \quad (3.22)$$

We can see that under the null hypothesis, we have $\mathbf{z} \sim N_q(\mathbf{0}, \mathbf{I}_q)$. The exact distribution of \mathbf{W} is complicated and is not tractable except for some special cases when k is small.

Taking 2-sample cases as an example, note that \mathbf{C} can be decomposed into two blocks of size $p \times p$ so that $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2]$ with \mathbf{C}_1 consisting of the first p

columns of \mathbf{C} , \mathbf{C}_2 the second p columns of \mathbf{C} , where $\mathbf{C}_1 = \mathbf{I}_p$ and $\mathbf{C}_2 = -\mathbf{I}_p$. Set

$$\mathbf{H}_l = (\mathbf{C}\Sigma_\beta\mathbf{C}^T)^{-1/2}\mathbf{C}_l, l = 1, 2, \text{ so that } \mathbf{H} = (\mathbf{C}\Sigma_\beta\mathbf{C}^T)^{-1/2}\mathbf{C} = (\mathbf{H}_1, \mathbf{H}_2).$$

We have

$$\mathbf{C}\Sigma_\beta\mathbf{C}^T = \sigma_1^2(\mathbf{X}_1^T\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2^T\mathbf{X}_2)^{-1} \quad (3.23)$$

and

$$\begin{aligned} \mathbf{H} &= (\mathbf{C}\Sigma_\beta\mathbf{C}^T)^{-1/2}\mathbf{C} = (\sigma_1^2(\mathbf{X}_1^T\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2^T\mathbf{X}_2)^{-1})^{-1/2}(\mathbf{I}_p, -\mathbf{I}_p) \\ &= ([\sigma_1^2(\mathbf{X}_1^T\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}]^{-1/2}, -[\sigma_1^2(\mathbf{X}_1^T\mathbf{X}_1)^{-1} + \sigma_2^2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}]^{-1/2}) \\ &= (\mathbf{H}_1, \mathbf{H}_2) \end{aligned} \quad (3.24)$$

It follows that $\mathbf{W} = \mathbf{H}\hat{\Sigma}_\beta\mathbf{H}^T = \sum_{l=1}^k \mathbf{W}_l$, where $\mathbf{W}_l = \hat{\sigma}_l^2\mathbf{H}_l(\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{H}_l^T, l = 1, 2$.

Since it is known that \mathbf{z} follows the standard normal distribution, our interest is to approximate the distribution of \mathbf{W} . We will derive the approximated distribution of \mathbf{W} for general k -sample cases through following theorems.

Theorem 3.1. *We have*

$$\mathbf{W} \stackrel{d}{=} \sum_{l=1}^k \mathbf{W}_l, \quad \mathbf{W}_l \stackrel{d}{=} \frac{\chi_{n_l-p}^2}{n_l - p} \mathbf{G}_l, \quad (3.25)$$

where $\mathbf{W}_l, l = 1, 2, \dots, k$ are independent and $\mathbf{G}_l = \sigma_l^2\mathbf{H}_l(\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{H}_l^T$. Furthermore,

$$\mathbf{E}(\mathbf{W}) = \sum_{l=1}^k \mathbf{G}_l = \mathbf{I}_q, \quad \mathbf{E}\text{tr}(\mathbf{W} - \mathbf{E}\mathbf{W})^2 = 2 \sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2). \quad (3.26)$$

Proof of Theorem 3.1 The assertions in (3.25) follow directly from the definitions of $\mathbf{W}_l, \hat{\sigma}_l^2$ and \mathbf{G}_l , together with the distributions of $\hat{\sigma}_l^2, l = 1, 2, \dots, k$ given in (3.14). We now show the assertions given in (3.26). Since \mathbf{W}_l are independent, we have

$$\begin{aligned}
\mathbf{E}\mathbf{W} &= \sum_{l=1}^k \sigma_l^2 \mathbf{H}_l (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \mathbf{H}_l^T \\
&= (\mathbf{C}\Sigma_\beta \mathbf{C}^T)^{-\frac{1}{2}} \sum_{l=1}^k \sigma_l^2 \mathbf{C}_l (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \mathbf{C}_l^T (\mathbf{C}\Sigma_\beta \mathbf{C}^T)^{-\frac{1}{2}} = \mathbf{I}_q, \\
\text{Etr}(\mathbf{W} - \mathbf{E}\mathbf{W})^2 &= \sum_{a=1}^q \sum_{b=1}^q \text{Var}(\mathbf{W}_{ab}) \\
&= \sum_{a=1}^q \sum_{b=1}^q \sum_{l=1}^k \frac{2}{n_l - p} \sigma_l^4 (\mathbf{H}_l (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \mathbf{H}_l^T)_{ab}^2 \\
&= 2 \sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2),
\end{aligned}$$

where \mathbf{W}_{ab} denotes the (a, b) -th entry of \mathbf{W} . The theorem is proved.

By the random expression of \mathbf{W} given in Theorem 3.1, we may approximate \mathbf{W} by a random variable $\mathbf{R} \stackrel{d}{=} \frac{\chi_d^2}{d} \mathbf{G}$ where the unknown parameters d and \mathbf{G} are determined via matching the first moments and the total variations of \mathbf{W} and \mathbf{R} . Let the total variation of a random matrix $\mathbf{X} = (x_{ij}) : m \times m$ be defined as $\text{Etr}(\mathbf{X} - \mathbf{E}\mathbf{X})^2 = \sum_{i=1}^m \sum_{j=1}^m \text{var}(x_{ij})$, i.e., the sum of the variances of all the entries of \mathbf{X} . Then, we solve the following two equations for d and \mathbf{G} :

$$\mathbf{E}(\mathbf{W}) = \mathbf{E}(\mathbf{R}), \quad \text{Etr}(\mathbf{W} - \mathbf{E}\mathbf{W})^2 = \text{Etr}(\mathbf{R} - \mathbf{E}\mathbf{R})^2. \quad (3.27)$$

The solution of (3.27) is given in Theorem 3.2 below.

Theorem 3.2. *We have*

$$\mathbf{G} = \mathbf{I}_q, \quad d = \frac{q}{\sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2)}. \quad (3.28)$$

Moreover, we have the following lower bound for d :

$$d \geq (n_{\min} - p) \quad (3.29)$$

where $n_{\min} = \min_{1 \leq l \leq k} n_l$ is the minimum sample size of the k regression models.

Proof of Theorem 3.2 Since $\mathbf{R} \stackrel{d}{=} \frac{\chi_d^2}{d} \mathbf{G}$, we have $E(\mathbf{R}) = \mathbf{G}$ and $E\text{tr}(\mathbf{R} - E\mathbf{R})^2 = \frac{2}{d} \text{tr}(\mathbf{G}^2)$. Since $E(\mathbf{W}) = \mathbf{I}_q$, we have $\mathbf{G} = \mathbf{I}_q$ and $E\text{tr}(\mathbf{R} - E\mathbf{R})^2 = \frac{2q}{d}$. By Theorem 3.1, we have $E\text{tr}[(\mathbf{W} - E\mathbf{W})^2] = 2 \sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2)$. This implies that $d = \frac{q}{\sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2)}$.

Let $\mathbf{A}_l = (\mathbf{X}_l^T \mathbf{X}_l)^{-1/2} \mathbf{H}_l^T : p \times q$. It follows that $\mathbf{G}_l = \mathbf{A}_l^T \mathbf{A}_l : q \times q$ and $\mathbf{A}_l \mathbf{A}_l^T : p \times p$ have the same nonzero eigenvalues. It is seen that (1) \mathbf{G}_l is nonnegative semi-definite; (2) all the eigenvalues of \mathbf{G}_l are nonnegative; and (3) \mathbf{G}_l has at most p nonzero eigenvalues. Let $\lambda_{lj}, j = 1, 2, \dots, p$ be the p largest eigenvalues of \mathbf{G}_l . Then, together with Theorem 3.1, we have $\mathbf{I}_q - \mathbf{G}_l = \sum_{i=1, i \neq l}^k \mathbf{G}_i$ is also nonnegative. This implies that all the eigenvalues of \mathbf{G}_l are less than 1. Therefore, $0 \leq \lambda_{lj} \leq 1, j = 1, 2, \dots, p$. We are now ready to find the lower and upper bounds of d .

We now find the lower bound of d . By the above result, we have $\text{tr}(\mathbf{G}_l^2) = \sum_{j=1}^p \lambda_{lj}^2 \leq \sum_{j=1}^p \lambda_{lj} = \text{tr}(\mathbf{G}_l)$. Thus, by Theorem 3.1, we have

$$\frac{q}{d} = \sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\mathbf{G}_l^2) \leq (n_{\min} - p)^{-1} \sum_{l=1}^k \text{tr}(\mathbf{G}_l) = \frac{q}{n_{\min} - p},$$

which says that the lower bound of d is $(n_{\min} - p)$. The theorem is proved.

Theorem 3.2 suggests that the null distribution of T may be approximated by $qF_{q,d}$. The approximation will be good when d is large. By (3.28), we also see that when n_{\min} becomes large, d generally increases; and when $n_{\min} \rightarrow \infty$, we have $d \rightarrow \infty$ so that $T \sim \chi_q^2$ asymptotically.

In real data application, the approximate degree of freedom d has to be estimated based on the data. A natural estimator of d is obtained via replacing $\mathbf{G}_l, l = 1, 2, \dots, k$ by their estimators:

$$\hat{\mathbf{G}}_l = \hat{\sigma}_l^2 \hat{\mathbf{H}}_l (\mathbf{X}_l^T \mathbf{X}_l)^{-1} \hat{\mathbf{H}}_l^T, \quad l = 1, 2, \dots, k,$$

where $\hat{\mathbf{H}}_l = (\mathbf{C} \hat{\Sigma}_\beta \mathbf{C}^T)^{-1/2} \mathbf{C}_l$. Thus, $\hat{d} = \frac{q}{\sum_{l=1}^k (n_l - p)^{-1} \text{tr}(\hat{\mathbf{G}}_l^2)}$. Notice that $\sum_{l=1}^k \hat{\mathbf{G}}_l = \mathbf{I}_q$ so that the range of d given in (3.29) is also the range of \hat{d} .

In summary, the ADF test can be conducted using the usual F -distribution since

$$T \sim qF_{q,\hat{d}} \text{ approximately.} \quad (3.30)$$

In other words, the critical value of the ADF test can be specified as $qF_{q,\hat{d}}(\alpha)$ for the nominal significance level α . We reject the null hypothesis in (3.22) when this critical value is exceeded by the observed test statistic T . The ADF test can also be conducted via computing the P-value based on the approximate null distribution specified in (3.30). Notice that when $U \sim F_{q,v}$, it has up to r finite moments where r is the largest integer such that $r < v/2$. To assure that the approximate null distribution of the ADF test, as specified in (3.30), has up to r finite moments, the

minimum sample size must satisfy

$$n_{\min} > p + 2r, \quad (3.31)$$

which is obtained via using the lower bound of d (and \hat{d} as well) given in (3.29). The required minimum sample size is then $p + 2r + 1$. In particular, to make the approximate null distribution of the ADF test have a finite first moment, each of the heteroscedastic regression models should have at least $p + 3$ observations. This is reasonable since for the l -th regression model, there are $p + 1$ parameters in β_l and σ_l^2 . Thus, to make $\hat{\sigma}_l^2$ have at least 3 degrees of freedom, the l -th sample should have at least $p + 3$ observations.

Notice that the d as defined in (3.28) is jointly determined by the sample sizes n_l and the underlying error variances $\sigma_l^2, l = 1, 2, \dots, k$. By the definition of d in (3.28), it is seen that there are two special cases in which the ADF test may not perform well. The first case is when the sample sizes $n_l, l = 1, 2, \dots, k$ are very different from each other with n_{\min} close to the required minimum sample size as suggested by (3.31). In this case, the value of d will be dominated by n_{\min} and hence it may not give a good representation to other samples. The second case is when all the sample sizes are close to or smaller than the required minimum sample size. In this latter case, the value of d will also be small, leading to a less accurate approximation to the null distribution of the ADF test.

Chapter 4

Simulation and Real Life Example

4.1 Simulation

In this section, we will investigate the performance of the proposed ADF test via comparisons of several test statistics. As mentioned previously, the traditional Chow's test does not perform well for heteroscedastic case, thereafter we have a few modified Chow's tests, for example, Toyoda's or Conerly and Mansfield's modified Chow's test. These tests are all aimed to improve the effectiveness Chow's test in heteroscedastic case. In particular, we will compare the Chow's test, the Conerly and Mansfield's modified Chow's test and the ADF test in the following simulation studies.

4.1.1 Simulation: two sample cases

As a first example of the effectiveness of the new ADF approach, we conduct simulation studies to compare three test statistics, i.e., Chow's test, modified Chow's test(M-Chow) and the ADF test, for 2-sample cases. Our simulation model can be referred to Moreno, Torres and Casella (2005) and it is designed as follows:

$$M : Y_i = X_i\beta_i + \epsilon_i, \epsilon_i \sim N(0, \sigma_i^2), i = 1, 2, \quad (4.1)$$

There are four cases as listed in table (4.1). For each situation we generate values of X_i , each entry of a $n_1 \times p$ matrix, from a standard normal distribution except for the first column all being 1. The values of vector β_1 are generated from standard normal distribution and β_2 is set to be $\beta_1 + \delta$, where δ is the tuning parameter of the difference between β_1 and β_2 . When $\delta = 0$, when $\beta_1 = \beta_2$, i.e., the null hypothesis is true. In this case, the null hypothesis of equal variance holds and therefore if we record the P -values of test statistics in simulation study it gives the empirical size of the tests. Similarly when $\delta > 0$, we will have the power of the tests. The σ_1^2 and σ_2^2 are calculated by $2/(1 + \rho)$ and $2\rho/(1 + \rho)$. We should notice that the parameter ρ is designed to adjust the heteroscedasticity. When $\rho = 1$, we have $\sigma_1 = \sigma_2$ with respect to homogeneity case. When $\rho \neq 1$, we have the data for heteroscedasticity case. After we generate values for X_i , β_i and σ_i , we can have values for Y_i according to formula (4.1). We then apply the three tests to the generated data and record their P -values. This process is repeated $N=10000$ times.

Table 4.1: Parameter configurations for simulations

	Homogeneity	Heteroscedasticity
H_0 true	$\rho = 1, \delta = 0$	$\rho = 0.1, 10, \delta = 0$
H_A true	$\rho = 1, \delta = 0.1, \dots, 0.4,$	$\rho = 0.1, 10, \delta = 0.1, \dots, 0.4.$

The empirical size (when $\delta = 0$) and powers (when $\delta > 0$) of the three tests are the proportions of rejecting the null hypothesis, i.e., when their P -values are less than the nominal significance level α . In all the simulations conducted, we used $\alpha = 5\%$ for simplicity.

The empirical sizes and powers of the three tests for testing the equivalence of coefficients, together with the associated tuning parameters, are presented in Table (4.2) - (4.4), with the number of covariates to be $p = 5, 10, 20$ respectively. The columns labeled with “ $\delta = 0$ ” present the empirical sizes of these tests, whereas the columns labeled with “ $\delta > 0$ ” show the power of the tests. To measure the overall performance of a test in terms of maintaining the nominal size α , we define the average relative error as

$$\text{ARE} = M^{-1} \sum_{j=1}^M |\hat{\alpha}_j - \alpha| / \alpha \times 100 \quad (4.2)$$

where $\hat{\alpha}_j$ denotes the j -th empirical size for $j = 1, 2, \dots, M$, $\alpha = 0.05$ and M is the number of empirical sizes under consideration. The smaller ARE value indicates that the better overall performance of the associated test. Conventionally, when

$ARE \leq 10$, the test performs very well; when $10 < ARE \leq 20$, the test performs reasonably well; and when $ARE > 20$, the test does not perform well since its empirical sizes are either too liberal or too conservative and therefore may be unacceptable. The ARE values of the three tests are also listed in the bottom of these three tables.

As a starting point, we compare the Chow's test, modified Chow's test and the ADF test by examining their empirical sizes which are listed in the columns labeled with $\delta = 0$. When the homogeneity assumption is valid, i.e., $\sigma_1 = \sigma_2$, the empirical sizes of three tests are comparable. Under heteroscedasticity, it can be observed that the range of the first column is 0.004 - 0.24, with a few values having large deviation from 0.05. The values of second and third columns are close and comparable, with smaller range 0.049 - 0.056 and 0.049 - 0.055 respectively, which have much smaller deviation from 0.05. Hence we may conclude that the modified Chow's test and ADF test perform better in maintaining the empirical size. Since the values of these two columns are close, it is not easy to decide the superior one. The ARE values of these two columns are 4.18 and 4, which suggest that the ADF test perform slightly better with smaller ARE value in terms of maintaining the empirical sizes. It also should be noticed that the empirical sizes of Chow's test with large deviation from 0.05 appears in the second and third block of variance specification, where the homogeneity condition no longer holds. It is in comply with the previous literature review, which concludes that the Chow's test does not

perform well in heteroscedastic cases and it is not stable in maintaining the empirical sizes. For $\delta > 0$ cases, we have the power of the tests listed in the table. The power of tests increases as δ increases. For homogeneous variances, these three tests perform comparably well with similar values of power. It does not make much sense to compare the power of Chow's test with those of modified Chow's test and the ADF test under heteroscedasticity since their empirical sizes are very different. Nevertheless we can compare the performance of modified Chow's test and ADF test under heteroscedasticity. The ADF test performs slightly better with larger power for heteroscedastic cases.

Table 4.2: Empirical sizes and powers for 2-sample test($p = 5$).

		$\delta = 0$			$\delta = 0.1$			$\delta = 0.2$			$\delta = 0.3$			$\delta = 0.4$		
(σ_1, σ_2)	n	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF
(1, 1)	(40, 40)	.052	.051	.051	.094	.093	.092	.251	.250	.249	.547	.546	.545	.820	.819	.818
	(50, 30)	.057	.056	.055	.091	.092	.091	.246	.240	.239	.516	.508	.507	.781	.770	.769
	(50, 90)	.050	.052	.052	.127	.123	.123	.428	.418	.417	.805	.797	.796	.973	.971	.971
(1.35, 0.43)	(40, 40)	.057	.049	.049	.104	.090	.093	.265	.239	.253	.549	.517	.543	.816	.793	.815
	(50, 30)	.010	.056	.055	.019	.092	.099	.093	.278	.298	.311	.603	.634	.622	.864	.885
	(50, 90)	.240	.051	.051	.376	.106	.109	.675	.331	.341	.909	.677	.689	.987	.914	.918
(0.43, 1.35)	(40, 40)	.062	.051	.053	.096	.081	.086	.260	.233	.248	.550	.518	.548	.812	.789	.811
	(50, 30)	.223	.050	.050	.276	.076	.078	.473	.181	.191	.725	.389	.415	.886	.649	.670
	(50, 90)	.004	.051	.050	.025	.146	.151	.202	.529	.554	.640	.898	.908	.944	.995	.996
ARE		106.13	4.18	4												

Table 4.3: Empirical sizes and powers for 2-sample test($p = 10$).

		$\delta = 0$			$\delta = 0.1$			$\delta = 0.2$			$\delta = 0.3$			$\delta = 0.4$		
(σ_1, σ_2)	n	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF
(1, 1)	(40, 40)	.051	.050	.049	.101	.099	.098	.330	.327	.321	.702	.699	.693	.931	.929	.928
	(50, 30)	.051	.054	.052	.101	.100	.097	.311	.294	.290	.662	.630	.625	.915	.891	.891
	(50, 90)	.050	.052	.051	.158	.151	.151	.595	.578	.575	.945	.935	.935	.998	.997	.997
(1.35, 0.43)	(40, 40)	.068	.052	.051	.118	.093	.101	.344	.290	.337	.682	.625	.687	.922	.898	.925
	(50, 30)	.006	.052	.050	.013	.106	.114	.092	.346	.388	.366	.738	.793	.747	.952	.969
	(50, 90)	.347	.049	.048	.540	.120	.130	.864	.432	.463	.985	.830	.851	1.00	.980	.982
(0.43, 1.35)	(40, 40)	.067	.052	.049	.124	.098	.102	.328	.280	.319	.686	.627	.690	.923	.894	.924
	(50, 30)	.337	.053	.052	.420	.083	.089	.651	.202	.230	.876	.456	.509	.974	.734	.780
	(50, 90)	.002	.049	.049	.020	.185	.199	.251	.705	.741	.807	.982	.987	.991	.999	1.00
ARE		158.29	4.02	2.49												

In Table (4.3), the range of empirical sizes of Chow's test, modified Chow's test and the ADF test are 0.006 - 0.347, 0.049 - 0.054 and 0.048 - 0.052 respectively. The corresponding ARE are 158.29, 4.02 and 2.49. Similar conclusion can be drawn from this table. The empirical sizes and powers for 2-sample case when $p = 20$ are listed in table (4.4) which also presents similar patterns which we have observed. It should be noticed that as p increases, the Chow's test perform even worse with larger deviation from 0.05 in maintaining empirical size under heteroscedasticity. Overall for 2-sample case, the Chow's test, modified Chow's test and ADF test perform comparably well under homogeneity. The ADF test is the best in maintaining empirical sizes with smallest ARE and it has larger powers than modified Chow's test when homogeneity assumption is no longer valid. Therefore we generally prefer

Table 4.4: Empirical sizes and powers for 2-sample test($p = 20$).

		$\delta = 0$			$\delta = 0.1$			$\delta = 0.2$			$\delta = 0.3$			$\delta = 0.4$		
(σ_1, σ_2)	n	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF
(1, 1)	(40, 40)	.052	.048	.045	.102	.098	.091	.337	.330	.310	.736	.727	.706	.958	.955	.949
	(50, 30)	.053	.060	.057	.097	.096	.083	.316	.268	.239	.697	.585	.558	.932	.852	.847
	(50, 90)	.046	.050	.048	.182	.172	.167	.736	.695	.690	.989	.982	.983	1.00	1.00	1.00
(1.35, 0.43)	(40, 40)	.088	.059	.052	.143	.098	.105	.368	.281	.350	.714	.606	.729	.938	.889	.951
	(50, 30)	.002	.052	.045	.007	.109	.106	.051	.364	.407	.272	.772	.837	.647	.967	.982
	(50, 90)	.568	.049	.050	.761	.128	.141	.972	.481	.553	1.00	.895	.931	1.00	.993	.997
(0.43, 1.35)	(40, 40)	.081	.053	.052	.132	.086	.096	.359	.269	.340	.724	.617	.737	.939	.888	.949
	(50, 30)	.590	.058	.057	.678	.077	.086	.852	.169	.205	.964	.342	.436	.996	.593	.715
	(50, 90)	.001	.050	.051	.012	.226	.249	.272	.835	.886	.885	.998	.999	.999	1.00	1.00
ARE		273.87	7.36	7												

the ADF test for comparing the coefficients of two linear regression models.

4.1.2 Simulation: multi-sample cases

In this section, we will compare the performance of four test statistics for k -sample case. These are Chow's test in (3.5), Wald-type test in (3.20), modified Chow's test in (3.9) and the ADF test in (3.30). First we will consider 3-sample case. The data generating procedures are similar to 2-sample case and the simulation results are listed in the table below. By observing the table, the first thing should be noticed is that the entries of Wald's test and Chow's test are equal. It can be justified because under the homogeneity assumption, we can prove the test

Table 4.5: Empirical sizes and powers for 3-sample test($p = 10$).

		$\delta = 0$				$\delta = 0.1$				$\delta = 0.2$				$\delta = 0.3$			
$(\sigma_1, \sigma_2, \sigma_3)$	\mathbf{n}	Wald	Chow	ADF	M-Chow	Wald	Chow	ADF	M-Chow	Wald	Chow	ADF	M-Chow	Wald	Chow	ADF	M-Chow
(1, 1, 1)	(60, 60, 60)	.066	.066	.052	.064	.313	.313	.275	.309	.962	.962	.947	.961	1.00	1.00	1.00	1.00
	(60, 45, 65)	.045	.045	.039	.043	.258	.258	.215	.252	.913	.913	.874	.908	1.00	1.00	0.998	1.00
	(75, 350, 100)	.042	.042	.036	.045	.801	.801	.762	.781	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(1.29, 0.58, 1)	(60, 60, 60)	.065	.065	.035	.049	.338	.338	.342	.295	.950	.950	.978	.939	1.00	1.00	1.00	1.00
	(60, 45, 65)	.033	.033	.045	.054	.194	.194	.304	.235	.865	.865	.955	.901	1.00	1.00	1.00	1.00
	(75, 350, 100)	.651	.651	.035	.048	.988	.988	.670	.627	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(0.58, 1.29, 1)	(60, 60, 60)	.070	.070	.052	.058	.323	.323	.340	.273	.955	.955	.975	.937	1.00	1.00	1.00	1.00
	(60, 45, 65)	.120	.120	.041	.056	.364	.364	.272	.226	.947	.947	.942	.861	1.00	1.00	1.00	1.00
	(75, 350, 100)	.001	.001	.041	.041	.257	.257	.948	.835	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ARE		178	178	18.22	12.44												

statistic in (3.5) is equivalent to that of (3.20). Therefore in the later simulation studies for more sample case, we will only record one result of these two test statistics. The ARE of these two statistics suggest that they performs badly in terms of maintaining the empirical size. Their empirical size are either too conservative or too liberal, ranging from 0.1% to 65.1%. On the other hand, the ARE of modified Chow's test is smaller than that of the ADF test, which indicates that the modified Chow's test performs better than the ADF test in terms of maintaining the empirical size. Since the empirical size of these four tests are very different, it does not make much sense to compare their power. We can still make comparisons of the power of ADF and modified Chow's tests. When $\delta = 0.1$ and 0.2 , the power of ADF test is larger than that of modified Chow-type test. When δ becomes larger, the powers of these two tests become comparable. Therefore in terms of the power, the ADF test performs better than the modified Chow-type

test and overall speaking, these two test performs much better than Chow's test and the modified Chow's test outperforms the ADF test in terms of maintaining empirical size. We then conduct the simulation studies for 5-sample case. There are two cases under consideration, i.e., $p = 2$ and $p = 5$ and the simulation results are listed in the tables below. These tables presents similar results as table (4.5). The smaller ARE of the modified Chow's test indicates its better performance in terms of maintaining the empirical size. Although the ARE of the ADF is larger, it is still acceptable because its values are around 20. The AREs of the Chow test are very large, with empirical size ranging from 1.3% to 31.9% in table (4.6) and from 1.2% to 22.7% in table (4.7). As for the power of the modified Chow's test and ADF test, they are comparable most of the time. Overall we will prefer the modified Chow's test due to its better performance in maintaining the empirical size.

From these tables above, we may conclude that for 2-sample cases, the new proposed ADF test is the most preferable test, since it performs the best in maintaining empirical sizes and has the largest power. For the generalized k -sample cases, we may choose the modified Chow's test because of its best performance in maintaining empirical sizes. The ADF test is comparable with slightly larger ARE values. The Chow's test is not acceptable for both cases because it is very unstable in maintaining empirical sizes.

Table 4.6: Empirical sizes and powers for 5-sample test($p = 2$).

		$\delta = 0$			$\delta = 0.1$			$\delta = 0.2$			$\delta = 0.3$		
$(\sigma_1, \dots, \sigma_5)$	n	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF	Chow	M-Chow	ADF
(1_5)	(15_5)	.047	.042	.042	.302	.285	.284	.894	.874	.874	.999	.998	.998
	$(15, 30_4)$.050	.047	.047	.325	.321	.327	.927	.921	.924	1.00	.999	.999
	$(30, 15_4)$.051	.046	.047	.562	.529	.539	.995	.993	.993	1.00	1.00	1.00
$(1_4, 2)$	(15_5)	.072	.044	.044	.201	.147	.148	.735	.610	.609	.975	.940	.941
	$(15, 30_4)$.053	.046	.046	.181	.156	.156	.763	.712	.711	.988	.982	.982
	$(30, 15_4)$.107	.061	.068	.418	.269	.282	.965	.884	.891	1.00	.995	.997
$(1_4, 4)$	(15_5)	.113	.042	.042	.160	.082	.082	.296	.138	.138	.645	.352	.353
	$(15, 30_4)$.079	.042	.042	.119	.076	.075	.301	.193	.190	.709	.504	.500
	$(30, 15_4)$.181	.063	.063	.276	.091	.096	.648	.256	.267	.964	.626	.643
$(2, 1_4)$	(15_5)	.074	.055	.055	.228	.173	.173	.660	.567	.567	.925	.888	.888
	$(15, 30_4)$.181	.048	.060	.368	.182	.198	.779	.560	.580	.961	.874	.889
	$(30, 15_4)$.017	.039	.037	.207	.318	.302	.749	.868	.857	.990	.999	.999
$(4, 1_4)$	(15_5)	.148	.070	.069	.198	.106	.106	.312	.182	.183	.555	.395	.395
	$(15, 30_4)$.319	.064	.075	.410	.094	.111	.595	.203	.228	.793	.345	.389
	$(30, 15_4)$.013	.043	.040	.054	.125	.115	.213	.369	.356	.518	.673	.663
ARE		120.13	16.53	20.40									

Table 4.7: Empirical sizes and powers for 5-sample test($p = 5$).

[illegible]

4.2 Real Life Examples

4.2.1 Example for 2-sample Case: *abundance of selected animal species*

MacPherson (1990, p. 513) described a study comparing two species of seaweed with different morphological characteristics. For each species of seaweed, the relationship between its biomass (dry weight) and the abundance of animal species that used the plant as a host, was measured. The data can be referred to Moreno et al. (2005, p.130). For each seaweed species, $\log(\text{abundance})$ is regressed on dry weight, and the question of interest is whether the relationship is the same in the two species. The following is the scatterplot of the data and the fitted least-squares lines. We apply the Chow's test, the modified Chow's test and the ADF test to the data set. The test statistics and p -values are listed in the table below. Moreno et al.

Table 4.8: Test Results

Test	Statistic	p -value
Modified Chow's Test	0.806	0.4535
ADF Test	0.812	0.4544

(2005) mentioned that the homogeneity assumption of these two linear models is questioned since the residual standard errors from individual regressions are 0.459 and 0.293 respectively. There is also evidence of heteroscedasticity in scatterplot (Fig 4.1). Therefore, the modified Chow's test and ADF test under heteroscedas-

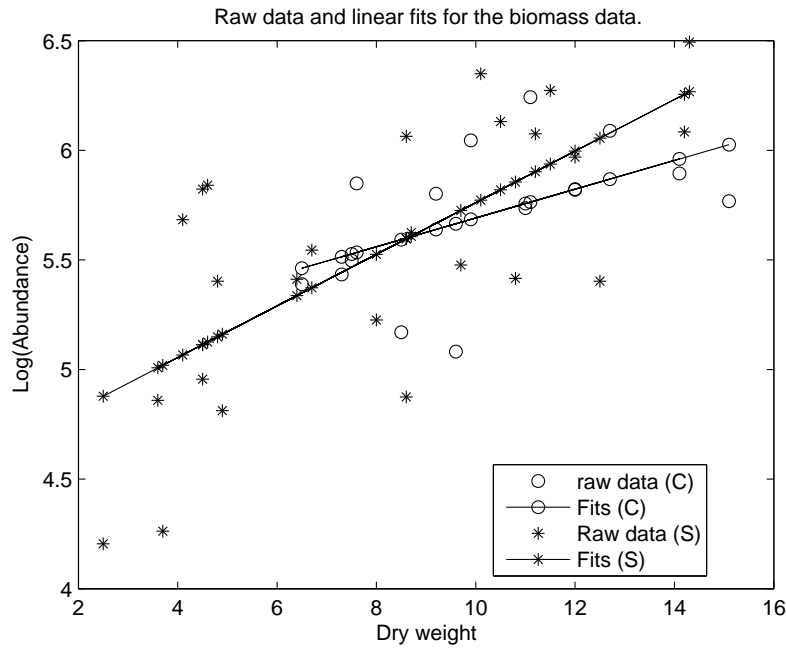


Figure 4.1: Scatter plot of data for Dry Weight vs. log(Abundance)

ticity are preferred. According to Moreno et al., a standard analysis, fitting a common regression with separate slopes and intercept yields a p -value of 0.0477 for the common intercept hypothesis and 0.0153 for the common slope hypothesis, which would lead to the conclusion that the animal species response is different. However, the p -value of modified Chow's test and ADF test are 0.4535 and 0.4544, which indicate that the null hypothesis can hardly be rejected. If the equivalence of coefficients can not be rejected, it suggests that similar relationships in the two species. Therefore, we may further conclude that there is evidence for similar animal species response when heteroscedasticity is accounted for. This conclusion is consistent with Moreno's results where they obtained a modified Chow statistic p -value of 0.634 and a posterior probability of H_0 of 0.997 using the intrinsic Bayes

factor.

4.2.2 Example for 10-sample Case : *investment of 10 large American corporations*

A classical model of investment demand is defined by

$$I_{it} = \alpha_i + \beta F_{it} + \gamma C_{it} + \epsilon_{it} \quad (4.3)$$

where i is the index of the firms, t is the time point, I is the gross investment, F is the market value of the firm and C is the value of the stock of plant and equipment. In this section we investigate the Grunfeld (1958) data by fitting model (4.3) and test the equivalence of coefficients. We are interested to analyze the relationship between the dependent variable I and explanatory variable F and C of 10 American corporation during the period of 1935 to 1954.

The test results are listed in table(4.9). The range of estimated standard errors

Table 4.9: Test Results				
Test	Statistic	df_1	df_2	p -value
Chow's Test	39.83	18	180	0
Modified Chow's Test	48.91	18	19.7	0
ADF Test	49.58	3.6	35.7	0

of ten linear models are 1.06 to 108.89, which strongly suggests heteroscedasticity.

When the homogeneity assumption no longer holds, we generally prefer the modified Chow's test and ADF test since they are more robust under heteroscedasticity. The p values of these two tests indicate that there is strong evidence to reject the null hypothesis of equivalent coefficients of the linear models. Therefore we may conclude that the investment pattern of these 10 American corporations are different.

Chapter 5

Conclusion and Future Research

This thesis introduces some new tools to test the coefficient of linear models for 2-sample and k -sample cases under the heteroscedasticity assumption. We generalize the traditional Chow's test to k -sample case under the homogeneity assumption. We also generalize the modified Chow's test to k -sample case by matching the moments of test statistics to a chi-square distribution. The Wald test can be easily generalized to k -sample case. Under the homogeneity assumption, the test statistic follows an F distribution. When homogeneity assumption no longer holds, we still use F random variable to approximate the test statistic, and follows some techniques to calculate the approximated degrees of freedom.

The simulation studies suggest that the generalized Chow's test performs badly in maintaining the empirical size and the modified Chow's test and Wald-type test perform much better. For 2-sample cases, the ADF test is preferred. Since the

modified Chow's test has smallest ARE and comparable power for k-sample cases, generally we prefer the Chow-type test when we test the equivalence of coefficient of linear models. The ADF test performs very well for 2-sample case, however, it seems not robust in k-sample case. The future research on the ADF test for k-sample case is guaranteed.

Chapter 6

Appendix

6.1 Matlab code in common for simulations

```
%-----
%Chow's test, modified Chow's test and New method for 2-sample cases
%-----

%% Data extraction

[n,p]=size(xy);

n1=gsize(1);n2=gsize(2);

X1=xy(1:n1,1:(p-1));y1=xy(1:n1,p);

X2=xy((n1+1):n,1:(p-1));y2=xy((n1+1):n,p);

p=p-1; %% dimension of X1, X2

%% Basic statistics computation

A1=inv(X1'*X1);
```

```

beta1=A1*X1'*y1;

hsigma21=y1'*(eye(n1)-X1*A1*X1')*y1/(n1-p);

A2=inv(X2'*X2);

beta2=A2*X2'*y2;

hsigma22=y2'*(eye(n2)-X2*A2*X2')*y2/(n2-p);

if method==0, %% Chow (1960) test, Homogeneity case

hsigma2=((n1-p)*hsigma21+(n2-p)*hsigma22)/(n-2*p); %%pooled variance

hsigma21=hsigma2;hsigma22=hsigma2;

Sigma=hsigma2*(A1+A2);

iSigma=inv(Sigma);

df1=p;df2=n-2*p;

stat=(beta1-beta2)'*iSigma*(beta1-beta2)/df1;


elseif method==1, %% Modified Chow method (Conerly and Mansfield 1988)

Heteroscedasticity case

W=X1'*X1*inv(X1'*X1+X2'*X2);

[U,D]=eig(W);

d=diag(D);

dbar=mean(d);

X=[X1;X2];H=X*inv(X'*X)*X';y=[y1;y2];

H1=X1*A1*X1';H2=X2*A2*X2';

```

```

temp0=y'*(eye(n)-H)*y-(n1-p)*hsigma21-(n2-p)*hsigma22;

temp1=(1-dbar)*hsigma21+dbar*hsigma22;

stat=temp0/temp1/p;

df1=(p*temp1)^2/sum(((1-d)*hsigma21+d*hsigma22).^2);

df2=temp1^2/(((1-dbar)*hsigma21)^2/(n1-p)+(dbar*hsigma22)^2/(n2-p));

elseif method==2, %% New method Heteroscedasticity case

Sigma=hsigma21*A1+hsigma22*A2;

iSigma=inv(Sigma);

G1=hsigma21*A1*iSigma;

G2=hsigma22*A2*iSigma;

df1=p;

df2=p/(trace(G1^2)/(n1-p)+trace(G2^2)/(n2-p));

stat=(beta1-beta2)'*iSigma*(beta1-beta2)/df1;

%% Computing the statistic

stat=(y'*Q*y)/(a11+a22);

%% Computing the dfs

df1=(a11+a22)^2/trace(H^2);

df2=(a11+a22)^2/(a11^2/(n1-p)+a22^2/(n2-p));

end

```

```

pvalue=1-fcdf(stat,df1,df2);

pstat=[stat,pvalue];

params=[df1,df2];

vbeta=[beta1,beta2];

vhsigma=[hsigma21,hsigma22];

%-----

Wald-type test and Wald-type ADF test for k-sample cases

%-----

%% Data extraction

k=length(gsize);

[N,q]=size(xy);

p=q-1; %% dimension of Xi

if nargin<3|length(method)==0,method=1;end

if nargin<5|length(c)==0,c=0;end

if nargin<4|length(C)==0,

C=[eye(p),zeros(p,(k-2)*p),-eye(p)];

for i=2:(k-1),

C=[C;[zeros(p,(i-1)*p),eye(p),zeros(p,(k-i-1)*p),-eye(p)]];

end

```

```

end

q=size(C,1);

%% Basic statistics computation

B=0;Sigma0=0;Sigma1=0;

for i=1:k, %% Fitting

ni=gsize(i);

if i==1,

temp=0;

else

temp=temp+gsize(i-1);

end

flagi=(temp+1):(temp+ni);

Xi=xy(flagi,1:p);yi=xy(flagi,p+1);

Di=inv(Xi'*Xi);

vhbeta(:,i)=Di*Xi'*yi;

vhsigma2(i)=sum((yi-Xi*vhbeta(:,i)).^2)/(ni-p);

Ci=C(:,((i-1)*p+1):(i*p));

B=B+Ci*vhbeta(:,i);

Sigma0=Sigma0+Ci*Di*Ci'; %% C*D*C'

if method==1,

```



```

Ai=Di*vhsigma2(i);

Sigma1=Sigma1+Ci*Ai*Ci'; %% C*Sigma(beta)*C'

end

end

if method==0, %% assume homogeneity

df1=q;df2=sum(gsize-p);

%% Computing the pooled variance

hsigma2=sum((gsize-p).*vhsigma2)/df2;

Sigma=Sigma0*hsigma2;

%% Computing the test statistic

iSigma=inv(Sigma);

stat=(B-c)'*iSigma*(B-c)/df1; %% Test statistic

%ihSigma=sqrtm(iSigma);

pvalue=1-fcdf(stat,df1,df2);

pstat=[stat,pvalue];

params=[df1,df2];

elseif method==1, % chi2-approximation matching 2-cumulants, F-type test

iSigma=inv(Sigma0);

Omega=Sigma1*iSigma;

temp0=trace(Omega^2);temp=trace(Omega);

```

```

beta1=temp0/temp;df1=temp^2/temp0;

temp0=sum(vhsigma2.^2.*(gsize-p));

temp=sum(vhsigma2.*(gsize-p));

beta2=temp0/temp;df2=temp^2/temp0;

%% Computing the test statistic

iSigma=inv(Sigma0);

stat0=(B-c)'*iSigma*(B-c)/temp; %% Test statistic

%ihSigma=sqrtm(iSigma);

stat=stat0*(beta2*df2)/(beta1*df1);

pvalue=1-fcdf(stat,df1,df2);

pstat=[stat,pvalue];

params=[df1,df2];

end

```

6.2 Simulation Studies for 2-sample and k-sample cases

```

%-----

Simulation for 2-sample cases for comparing Chow test, modified Chow test
and new test

%-----

```

```

%%Simulation parameter configurations

Vrho=[1/10,1,10];

nVrho=size(Vrho,2);

sigma20=2;

Gsize=[40,40;

       50,30;

       50,90];

nGsize=size(Gsize,1);

Vp=[5,10,20];

nVp=size(Vp,2);

for i5=1:nVp,

p=Vp(i5);

Vdelta=[0,.1,.2,.3,.4,.5,.6,.7,.8]/2;

nVdelta=length(Vdelta);

nsim=10000;alpha=.05;

disp(['Number of replicates=',num2str(nsim)])

disp(['alpha=',num2str(alpha)])

disp('Chow, Modified Chow, New test')

disp('Empirical Powers')

time0=clock;

for iv=1:nGsize, %% specify the sample size

```

```

gsize=Gsize(iv,:);

n1=gsize(1);n2=gsize(2);

disp('sample sizes')

disp(gsize)

vpw=[];

for iii=1:nVrho, %% Specify the std

rho=Vrho(iii);

sigma1=sqrt(sigma20/(1+rho));

sigma2=sqrt(rho*sigma20/(1+rho));

%disp('sample std')

disp(['rho=',num2str(rho),', p=',num2str(p), ',

[n1,n2]=[' ,num2str(n1),',',num2str(n2),']'])

disp(['[sigma1,sigma2]=[' ,num2str(sigma1),',',num2str(sigma2),']'])

for ii=1:nVdelta, %% specify delta

delta=Vdelta(ii);

beta1=randn(p,1);

beta2=beta1+delta;

% disp(['delta=', num2str(delta)])

for i=1:nsim, %% simulation

%% data generating

X1=[ones(n1,1),randn(n1,p-1)];

```

```

X2=[ones(n2,1),randn(n2,p-1)];

y1=X1*beta1+randn(n1,1)*sigma1;

xy1=[X1,y1];

y2=X2*beta2+randn(n2,1)*sigma2;

xy2=[X2,y2];

xy=[xy1;xy2];

%% Testing

[pstat0,param0]=coefBF(xy,gsize,0); %% Chow test

[pstat1,param1]=coefBF(xy,gsize,1); %% Modified Chow test

[pstat2,param2]=coefBF(xy,gsize,2); %% New test

vpval(i,:)= [pstat0(2),pstat1(2),pstat2(2)];

vparam(i,:)= [param0(2),param1(2),param2(2)];

end %% end for i

pw(ii,:)=mean(vpval<alpha); %% Rate of rejections

mdf=mean(vparam);

disp([ii,pw(ii,:),mdf])

end %% for ii

vpw=[vpw,pw];

end %% for iii

disp(Vdelta)

disp(Vrho)

```

```

disp('empirical power')

disp(vpw)

end %% for iv

%-----

Simulation for 3-sample cases for comparing Chow-type and Wald-type
tests

%-----

%%Simulation parameter configurations

Vrho=[1/5,1,5];

nVrho=size(Vrho,2);

sigma20=2;

Gsize=[12,12,12;

       12,9,13;

       15,70,20]*5;

nGsize=size(Gsize,1);

Vp=[5,10,20];

nVp=size(Vp,2);

p=10;

Vdelta=[0:.1:.3];

nVdelta=length(Vdelta);

nsim=1000;alpha=.05;

```

```

disp(['Number of replicates=',num2str(nsim)])

disp(['alpha=',num2str(alpha)])

for iv=1:nGsize, %% specify the sample size
    gsize=Gsize(iv,:);

    n1=gsiz(1);n2=gsiz(2);n3=gsiz(3);

    disp('sample sizes')

    disp(gsize)

    vpw=[];

    for iii=1:nVrho, %% Specify the std
        rho=Vrho(iii);

        sigma1=sqrt(sigma20/(1+rho));

        sigma2=sqrt(rho*sigma20/(1+rho));

        sigma3=1;

        %[sigma1,sigma2,sigma3].^2

        %disp('sample std')

        disp(['rho=',num2str(rho)])

        for ii=1:nVdelta, %% specify delta
            delta=Vdelta(ii);

            beta1=randn(p,1);

            beta2=beta1+delta;

            beta3=beta1+delta/2;

```

```

% disp(['delta=', num2str(delta)])

for i=1:nsim, %% simulation

%% data generating

X1=[ones(n1,1),rand(n1,p-1)];
X2=[ones(n2,1),rand(n2,p-1)];
X3=[ones(n3,1),rand(n3,p-1)];

y1=X1*beta1+randn(n1,1)*sigma1;
xy1=[X1,y1];

y2=X2*beta2+randn(n2,1)*sigma2;
xy2=[X2,y2];

y3=X3*beta3+randn(n3,1)*sigma3;
xy3=[X3,y3];

xy=[xy1;xy2;xy3];

%% Testing

[pstat10,param10]=coefBFk(xy,gsize,0);

%% Wald-type stat, F-test, homogeneity

[pstat11,param11]=coefBFk(xy,gsize,1);

%% Wald-type stat, ADF test, heteroscedasticity

[pstat20,param20]=coefkBF(xy,gsize,0);

%% Chow-type test, F-test, homogeneity

[pstat21,param21]=coefkBF(xy,gsize,1);

```



```

%% Modified Chow-type test, ADF test, heteroscedasticity
vpval(i,:)=pstat10(2),pstat20(2),pstat11(2),pstat21(2)];
vparam(i,:)=param10(2),param20(2),param11(2),param21(2)];
end %% end for i

pw(ii,:)=mean(vpval<alpha); %% Rate of rejections
mdf=mean(vparam);

disp([ii,pw(ii,:),mdf])

end %% for ii

vpw=[vpw,pw];

end %% for iii

disp(['p=',num2str(p)])

disp(['delta=',num2str(Vdelta)])

disp(['rho=',num2str(Vrho)])

disp('empirical power')

disp(vpw)

end %% for iv

```

Bibliography

- [1] M.M. Ali and J.L. Silver. Tests for equality between sets of coefficients in two linear regressions under heteroscedasticity. *Journal of the American Statistical Association*, 80(391):730–735, 1985.
- [2] A. Buse and N.K. Dastoor. AN HISTORICAL NOTE ON THE JAYATISSA TEST*. *The Manchester School*, 58(3):260–264, 1990.
- [3] G.C. Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.
- [4] M.D. Conerly and E.R. Mansfield. An approximate test for comparing heteroscedastic regression models. *Journal of the American Statistical Association*, 83(403):811–817, 1988.
- [5] M.D. Conerly and E.R. Mansfield. An approximate test for comparing independent regression models with unequal error variances* 1. *Journal of econometrics*, 40(2):239–259, 1989.

- [6] F.M. Fisher. Tests of equality between sets of coefficients in two linear regressions: an expository note. *Econometrica: Journal of the Econometric Society*, pages 361–366, 1970.
- [7] G. Ghilagaber. Another look at Chow’s test for the equality of two heteroscedastic regression models. *Quality & quantity*, 38(1):81–93, 2004.
- [8] SA Gupta. Testing the Equality Between Sets of Coefficients in Two Linear Regressions When Disturbances Are Unequal. *unpublished Ph. D. dissertation, Purdue University, Dept. of Statistics*, 1978.
- [9] Y. Honda and H. OHTANI. Modified Wald Tests in Tests of Equality between Sets of Coefficients in Two Linear Regressions under Heteroscedasticity. *The Manchester School*, 54(2):208–218, 1986.
- [10] JP Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, 1961.
- [11] WA Jayatissa. Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Econometrica*, 45(5):1291–92, 1977.
- [12] M. Kobayashi. A bounds test of equality between sets of coefficients in two linear regressions when disturbance variances are unequal. *Journal of the American Statistical Association*, 81(394):510–513, 1986.

- [13] E. Moreno, F. Torres, and G. Casella. Testing equality of regression coefficients in heteroscedastic normal regression models. *Journal of statistical planning and inference*, 131(1):117–134, 2005.
- [14] K. Ohtani and M. Kobayashi. A bounds test for equality between sets of coefficients in two linear regression models under heteroscedasticity. *Econometric Theory*, 2(02):220–231, 1986.
- [15] K. Ohtani and T. Toyoda. A monte carlo study of the wald, lm, and lr tests in a heteroscedastic linear model. *Communications in Statistics-Simulation and Computation*, 14(3):735–746, 1985.
- [16] K. Ohtani and T. Toyoda. Small sample properties of tests of equality between sets of coefficients in two linear regressions under heteroscedasticity. *International Economic Review*, 26(1):37–44, 1985.
- [17] P. Schmidt and R. Sickles. Some further evidence on the use of the Chow test under heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 1293–1298, 1977.
- [18] J.G. Thursby. A comparison of several exact and approximate tests for structural shift under heteroscedasticity. *Journal of Econometrics*, 53(1-3):363–386, 1992.
- [19] T. Toyoda. Use of the Chow test under heteroscedasticity. *Econometrica: Journal of the Econometric Society*, pages 601–608, 1974.

- [20] T. Toyoda and K. Ohtani. Testing equality between sets of coefficients after a preliminary test for equality of disturbance variances in two linear regressions* 1. *Journal of Econometrics*, 31(1):67–80, 1986.
- [21] P.A. Watt. TESTS OF EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS WHEN DISTURBANCE VARIANCES ARE UNEQUAL: SOME SMALL SAMPLE PROPERTIES*. *The Manchester School*, 47(4):391–396, 1979.
- [22] S. Weerahandi. Testing regression equality with unequal variances. *Econometrica: Journal of the Econometric Society*, pages 1211–1215, 1987.
- [23] A. Zellner. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368, 1962.